DOCUMENT RESUME

ED 050 158 TM 000 550

AUTHOR Sulkin, Howard A.; And Others
TITLE Problems and Techniques in Achieving Both Ecological and Population Validity in Educational Research.
INSTITUTION Chicago Univ., Ill. Industrial Relations Center.
PUB DATE Feb 71
NOTE 14p.; Paper presented at the Annual Meeting of the American Educational Research Association, New York, New York, February 1971

EDRS PRICE EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS Data Collection, *Evaluation, Interpersonal Relationship, Models, Research Design, *Research Problems, *Sampling, School Personnel, Statistical Analysis, *Validity
IDENTIFIERS Ecological Validity, Population Validity

ABSTRACT

Some of the requirements and consequences of rigorous and valid educational evaluation research are explored in terms of problems in achieving two types of external validity, population and ecological. The former refers to the generalizability of inferences to subjects not included in a study, while the latter is concerned with the "environment" under which the same results can be expected. A research model which emphasizes the use of a well-controlled and well-defined stimulus situation and thus facilitates unambiguous determination of the relationship between stimulus and response is considered. A recent study in which some social-psychological problems arose directly related to constraints involved in achieving ecological and population validity is examined in detail. (CK)

ED050158

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

# PROBLEMS AND TECHNIQUES IN ACHIEVING BOTH ECOLOGICAL AND POPULATION VALIDITY IN EDUCATIONAL RESEARCH

Howard A. Sulkin
Francis A. Chvatal
James R. Murray

Industrial Relations Center
University of Chicago

TM 000 550

Paper Presented at the American Educational Research Association
Meetings, New York
February, 1971

The purposes of this paper are to explore some of the requirements and consequences of rigorous and valid educational evaluation research: in terms of problems encountered in achieving, in particular, two types of external validity--population validity and ecological validity; to elaborate a model of such problems; and to examine a recent study in which some social-psychological problems arose which seem to be directly related to the constraints involved in achieving ecological and population validity.

A research model which has been entertained as an appropriate benchmark for educational evaluation research methodology is that of classical experimental psychology. Such a research model emphasizes the use of a stimulus situation which is both well-controlled and well-defined. The purpose of this research model is to facilitate unambiguous determination of the relationship between stimulus and response.

The complete appropriateness of such a research model in educational evaluation can be questioned. This experimental model can lead one to ignore whole classes of critical events in the research endeavor. To be specific, the quality of summative evaluations, which seek to determine the effectiveness of curriculum programs as actually used by teachers in the schools, can be affected by teacher/investigator interactions. Such interactions are often mediated by differences among schools which effect teachers' behavior, and these variables cannot be strictly controlled.

Bracht and Glass (1968) have presented two classes of threats to the external validity of experiments, referred to as population and ecological validity, as additions to the list formulated by Campbell and Stanley (1963). These two sources of error are especially relevant to the

problems of summative evaluation. Bracht and Glass specifically mention that knowledge of the effects of educational activities under natural conditions are in many instances of greater practical importance than the exactness of our knowledge of instruction-learning relationships in controlled situations. So that, at least from this point of view, it would be more valuable to achieve ecological and population validity in an evaluation project than to maintain a degree of experimental control consistent with the classical experimental model described earlier. It appears, however, that in achieving both of these types of validity yet another threat to external validity may be encountered. Before elaborating on this possible "new" threat, the concepts of population and ecological validity will be briefly reviewed.

Population validity refers to the generalizability of inferences with respect to subjects not included in the study. The fundamental problem here concerns the use of only a sample of subjects to make inferences about parameters of a population. Bracht and Glass focus on two threats to population validity. First, an investigator may confuse an experimentally accessible population with the target population. Second, one may be unaware of an interaction between personological variables and treatments. That is, some subject variable may interact with the experimental treatment to produce an effect which is not generalizable to all Ss.

These two threats to population validity, if ignored in research design, limit the generality of the inferences that can be drawn from the data. That is, regarding the first threat, one can generalize conclusions with statistical rigor only to that group of subjects that were able to be selected for inclusion in the experiment.

Some random sampling from a defined population is conceived to

be the most appropriate method for insuring a measure of generalizability of the findings. The second threat, the aptitude or subject by treatment interaction, is perhaps best avoided by stratifying the random sample on as many subject factors as feasible and judged potentially relevant to the treatments.

Ecological validity is the second class of threats suggested by Bracht and Glass. This concept refers to the "environment" of the experimental treatments. That is, under what settings, treatments, experimenters, response measures, etc., can the same results be expected? Ecological validity involves generalizing inferences over conditions (environments) other than those immediately involved in one's research. The conditions under which the research is executed must be as similar as possible to the conditions to which the research inferences are to be generalized. We are inclined to call this "natural setting research" to stress that data must be collected in a context which is typical of the environment in which the experimental variables will normally be applied.

This class of threats includes a number of specific problems; the Hawthorne effect, the Novelty effect, the Experimenter effect, as well as many others. The underlying theme here seems to involve walking the thin line between flexibility of treatment conditions on the one hand, and not allowing nontreatment variables to become confounded with treatment variables, on the other hand.

These two sources of invalidity are especially relevant to educational evaluation. It is of some interest, therefore, to examine the requirements of simultaneously achieving both types of validity in an evaluation study. First, students must be included in the evaluation in such a way that they can be taken to adequately "represent" the majority of stu-

dents who will probably use the materials or procedures being studied, i.e. population validity. A random sampling procedure, preferably using school districts as the primary sampling unit, would seem appropriate. This sampling plan should also include specific stratification or measurement procedures so that the possibility of treatment-by-personological variable interaction can be studied. Such variables as grade, social class, aptitude level, etc. are prime examples of such personological factors. The significant practical aspect of the procedures necessary to achieve population validity involves the unavoidable difficulties encountered in obtaining the participation of a sizeable number of randomly chosen, and therefore unfamiliar, schools and classrooms. Since schools differ in their resources and modes of operation, a random sampling procedure will force the investigator to cope with such differences among schools in order to achieve his objectives. This aspect of an evaluation can prove problematic for both the school and the researcher and is one basis for an additional threat to experimental validity.

Second, the situational or "conditional" integrity of the natural teaching environment should be maintained for any evaluation of curriculum materials, i.e. ecological validity. The goal of program evaluation in terms of this consideration is the measurement of the effects of curriculum materials as used in a normal and undisrupted teaching environment. It seems clear that in experimental research full achievement of such normality is near impossible. The nature of certain "effects", e.g. Hawthorne or Experimenter or Placebo, can be controlled but not eliminated. In attaining such control, it is necessary that an investigator succeeds in bringing about teacher understanding and consent with respect to usage of curriculum materials and corresponding measurement procedures, e.g. pre and postests. The process of

achieving such consent and understanding across randomly chosen schools and classrooms presents many problems which can threaten the validity of the inter-school treatment comparisons.

Thus, an attempt to satisfy both of these criterion of validity requires significant managerial and interpersonal skill on the part of the investigator if he is to successfully deal with these problems. Not only must a variety of distinct schools be contacted and recruited into the research, but the design requirements must be achieved equally well across these quite diverse educational settings. It is in terms of this implementation problem that we suggest a type of validity threat, the "organization effect", which seems distinctive enough to be considered in its own right.

The concept of validity entails a correspondence between the inferences and propositions made by the investigator concerning an experiment and the actual events in the experiment. From this perspective, the events which occur in the classroom related to curriculum usage comprise the phenomenon about which the investigator intends to make inferences. The difficulties mentioned above in terms of the attainment of population and ecological validity in an educational evaluation study arise because of the investigator's need to control and know about the actual usage of curriculum materials in the classrooms. Satisfaction of this research goal depends on the success the investigator has in working with the school personnel in terms of their behavior relevant to the use of the materials being evaluated.

There are two major aspects, each with three components, of the process of implementing an evaluation study. First, the verbal consent of both principals and teachers to the research must be obtained. Second, actual teacher performance in accordance with the study design must be achieved. Each of these aspects has, at least, three components. First,

the measurement procedures must be consented to. Second, the randomization pattern, which specifies the nature of the activities of teachers, must be accepted. Third, the intended or "correct" usage of the curriculum materials must be achieved. Consent and performance in terms of these three components of an evaluation are necessary if the objectives of the investigation are to be achieved.

The nature of the difficulty of successfully obtaining such commitment on the part of the school personnel appears to be similar to the problems discussed in terms of the social-psychology of experiments. The role demands of the experimenter-subject relationship have been used to explain various aspects of the outcomes of an experimental procedure. Argyris (1968) used the dimensions of organizational analysis in order to explain the distortions or transformations of experimental procedures which can occur because of the E-S relationship. He makes the point that the experimental situation can be regarded as a temporary organization and thereby subject to such an analysis. The major qualities which are generally taken to be relevant in such an analysis are: the degree of control which the E must exercise over the Ss; the effects which various control/participation relationships have on Ss; the motivations and goals of the Ss; the degree of time and energy which the Ss must contribute to the research; and the social context in which the research occurs from the Ss point of view. The temporary organization analogy focuses attention on the similarity between the man-boss relationship and the subject-experimenter relationship. The significance of this analogy resides in the propositions regarded as true concerning the man-boss relationship:

1. Highly authoritarian, i.e., one sided, relationships between the man (submissive) and the boss (dominant) leads to subordinate hostility, withdrawal and uncooperativeness;

2. Conversely, shared authority, responsibility and power between man and boss ameliorates such negative effects;
3. Differences in motivation and goals between a man and boss lead to differential commitment and performance on tasks;
4. Conversely, shared motivation and goals between man and boss leads to optimal performance and mutual commitment to task accomplishment.

Simple substitution of "subject-experimenter" for "man-boss" yields relevant statements for the research setting.

This organizational model of the experimental situation gains in significance when applied to educational evaluation since schools, which are actual organizations, are involved in the role of participants. If the principal and teachers in each school participating in an evaluation are not fully committed to their assigned duties in the evaluation, research success appears impossible. The capacity which the "subject" has to influence laboratory research is also possessed by school personnel in evaluation research. The quality of teaching can clearly be affected by teachers' knowledge of participating in an evaluation, e. g. his students being measured, as well as by the teachers' own values and beliefs about effective teaching methods which can affect curriculum usage. An additional factor in determining the kind of participation in an evaluation is the role played by the school principal. Each teacher will in fact determine the events in his own classroom. If the principal commits an unenthusiastic teacher to a certain course of action, e. g. using one of the programs being studies, the teacher can exhibit superficial cooperation but easily do less than a desirable job in using the materials.

There are some critical aspects of the investigator-principal-teacher relationships. Optimal performance requires that each teacher be committed to participate in the activity necessary for satisfying the

evaluation design. The role of the school principal is relevant in two respects. First, the principal can unilaterally commit teachers to participation, resulting in the possibility of teacher resentment, uncooperativeness or hostility. Second, a principal can make successful implementation difficult for committed teachers because of his own lack of commitment. In either case, the principal can play a role which is threatening to the success of the evaluation. The teachers, somewhat independently of the behavior of the principal, can bring about invalid research. Proper curriculum usage depends on teacher understanding and if teachers are unwilling to ask for clarification or help when problems are encountered, serious errors can be introduced into the evaluation process. Such threats to successful evaluation we suggest can be referred to as the "Organization Effect."

Argyris' analysis of the social-psychological nature of problems inherent in achieving validity and rigor in a research situation, particularly the effect of excessive control, suggest several specific strategies for dealing with these problems in an educational setting. These strategies include the development of cooperation, the sharing of responsibility; providing support, and responding to individual needs.

Developing Cooperation

An initial goal in such a strategy would be to develop cooperation between the teachers, administrators and investigators in phases of increasing mutual commitment. Much of this development of cooperation would be based upon successful clarification of roles, responsibilities and expectations of, and with, each of the people participating in the study. The investigator must come to understand the realistic pace at which this clarification can occur. A full conception of mutual roles and responsibilities cannot be understood and accepted from the beginning by the research

participant. This understanding is gradually developed. Only when this understanding is worked on in steadily increasing ways, can full and satisfactory cooperation really occur.

Shared Responsibility

The second aspect of the strategy is to develop an attitude of shared responsibility between the investigators and the teaching staff, i.e., research participants. Does the teaching staff understand the goals of the study, and do they see themselves having an opportunity to contribute to these goals? An immediate reaction of an investigator to this strategy may be a fear of the possibility of research contamination, but as Argyris (1968) concludes:

> Contamination. . . is inevitable. The issue therefore is not contamination versus no contamination. The issue is under what conditions can the researcher have the greatest awareness of, and control over, the degree of contamination.

It is believed that this approach to shared responsibility will help the investigator maintain control over the research contamination.

Another aspect of sharing responsibility involves "ownership" of the results. Those people cooperating in the study need to feel that they have not just been manipulated into the giving over of some information, but instead, that they are sharing equally in not only the formulation of the study, but in the outcome as well. Thus, feedback of results to participants would appear a wise policy.

Support

The third aspect of the strategy is to provide strong support to the teachers. It is particularly important to orient the teachers to

what they're expected to do in the educational research study, but it is also important to clarify what they can and cannot expect in support from the investigators throughout the length of the study. Follow-up with each teacher is particularly important because a person may have need of help, but for various reasons the person may be reluctant to ask for it.

Individualized Response

The fourth part of the strategy is closely related to step three above. Here, it is important to respond to field needs on an individualized basis. This response is particularly important when large numbers of classrooms are involved in a study. These classes are taught by teachers of varying skills, orientations, and needs. Unless these needs are seen by the teacher as being responded to sincerely and individually, an overt or covert hostile response is likely to occur.

In an evaluation study which attempted to achieve both ecological and population validity, many of the above problems were observed. Murray (1971) describes the research design and procedures used. Some of the events which occurred during that study give empirical support to some of the concepts discussed above. Briefly, 21 school districts were randomly selected from a sampling frame of 250 districts which were divided into three SES strata. A total of 124 classrooms participated in the evaluation study.

The role of the principal or coordinating administrator in these school districts proved to be problematic in six of the 21 districts. In another district, not counted as one of these 21, the school administrator actually led to the district being dropped from the research. Although letters and numerous phone conversations served as a basis for insuring mutual understanding and agreement, a number of serious errors occurred. In themselves, most of these errors were correctable.

The disturbing aspect of these events, however, is the possibility that they represent only a proportion, perhaps even a small one, of the deviations and errors made in the execution of the study. Obviously, these other "errors" are not known about.

In three of these cases, the school principal made commitments to participate in the curriculum evaluation without consultation with the appropriate teachers. Although the research staff held orientations with the teachers in order to gain voluntary commitment, the evidence at the close of the study was clear. Some of the teachers in these three schools verbally consented to participate but in fact never used the materials, although their principals repeatedly reported that the teachers would and were participating. Apparently, the principals were more concerned with their own goals and public image of commitment to research than to the actual feelings and motivations of their teachers.

In the remaining three districts, the principals were more the focus of problems than perhaps the cause. Repeated attempts to schedule pre and postests resulted in confusion at the time of measurement. Teachers reported that they had not been informed of the scheduled testing, or had not received the materials which had been sent in advance. It is not clear what the actual reasons for the confusion were, but in each case the administrator had assured the field staff that all was ready for the procedure as planned. In one case, in fact, classes which had not even been pretested were scheduled by the administrator for postesting!

The other major source of difficulties are the teachers themselves. In a number of cases, teachers who had committed themselves to participation in the study, without any indication of coercion, did not in fact use the assigned materials. In one school, two teachers who had been assigned different materials, decided on their own midway through the study, to switch materials. The results of such events, of

course, render that data unusable. Children had, in fact, had both sets of materials! For those teachers who did not use the materials assigned to them, many explanations are certainly possible. From a rigorous perspective, however, the possibility of an unobservable interaction between teacher usage and program effectiveness is a threat to basic inferences about program effectiveness.

In cases of administrative or teacher difficulties, the concepts of shared authority, responsibility and participant motives appear to provide useful hypotheses concerning such behavior. Efforts put forth, in the study reported here, were regarded by the investigators as substantially consistent with Argyris' suggestions. Yet, incidents such as those reported occurred nevertheless. The strength of organizational and behavioral styles in the schools appear to be such that a study must expect to have a certain percentage of failures or be prepared to invest significant energy into coordination and follow-up with the schools. In either case, many events in the schools go unknown to the investigator, and the possible effects on the validity of program evaluation may be significant.

## References


Argyris, C. "Some Unintended Consequences of Rigorous Research." Psychological Bulletin, 70, 1968, 185-197.

Bracht, G. H. and Glass, G. V. "The External Validity of Experiments." AERJ, 5, 1968, 437-474.

Campbell, D. T. and Stanley, J. C. "Experimental and Quasi-Experimental Designs for Research on Teaching." In N. L. Gage (ed.), Handbook of Research on Teaching. Chicago: Rand McNally and Co. 1963.

Murray, J. R. "An Experimental Design for the Summative Evaluation of Proprietory Reading Materials." Paper presented at AERA Meetings, New York, February, 1971.